I 21341

DOCUMENT RESUME

ED 055 733 RE 003 827

AUTHOR	Guba, Egon G.; Stufflebeam, Daniel L.
TITLE	Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action.
INSTITUTION	Indiana Univ., Bloomington. Measurement and Evaluation Center in Reading Education.
PUB DATE	Jun 70
NOTE	36p.
AVAILABLE FROM	Reading Program, School of Education, Indiana University, 202 Pine Hall, Bloomington, Indiana 47401, ($1.50)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Data Collection; *Decision Making; *Evaluation Criteria; *Evaluation Methods; *Evaluation Techniques; Program Evaluation; *Research Methodology

ABSTRACT

Part 1 of this monograph discusses the status of educational evaluation and describes several problems in carrying out such evaluation: (1) defining the educational setting, (2) defining decision types, (3) designing educational evaluation, (4) designing evaluation systems, and (5) defining criteria for judging evaluation. Part 2 proposes an approach to educational evaluation which would alleviate problems of definition and design. Different types of decision settings and appropriate evaluation strategies are explored. Types of decisions are matched with types of evaluation. The structure of evaluation design including the collection, organization, analysis, and reporting of information is outlined. Finally, criteria for judging evaluations such as validity, reliability, significance, scope, and efficiency are presented. Figures are given, including the CIPP (context, input, process, and product) Evaluation Model. (AL)

Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action

Egon G. Guba
Indiana University

Daniel L. Stufflebeam
Ohio State University

ED055733

Monograph Series in Reading Education, Indiana University, Bloomington

ED055733

Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action

Egon G. Guba
Indiana University

Daniel L. Stufflebeam
Ohio State University

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RE003 827

Monograph Series in Reading Education
Indiana University
Number 1, June 1970

Carl B. Smith
Series Editor

ED022123

This monograph was supported by the Measurement and Evaluation Center in Reading Education (MECRE) and the Laboratory for Educational Development, Indiana University, Bloomington, Indiana.

Preface

Evaluation has become a key concept and an essential operation in education today. Without a comprehensive evaluation system and techniques to provide a continuous monitoring of all educational activity, it is unreal for education to speak of accountability. For that reason, educators need to examine the theories and the research that propose ways to evaluate educational programs. This paper by Egon G. Guba and Daniel L. Stufflebeam was an important recent contribution to the discussion of how to develop a system for evaluating educational programs.

In using this monograph during training seminars and conferences for evaluators of reading programs, the Measurement and Evaluation Center in Reading Education found the participants received considerable benefit from reading and discussing the concepts presented. Many of the seminar participants suggested that the monograph be published for wide distribution.

Thus, this paper was placed in the Monograph in Reading series as a service to specialists in reading and as a service to the wider audience of evaluators who will find the suggestions contained herein helpful in developing further theory and practice in educational evaluation.

Carl B. Smith
Indiana University
April 1970

Acknowledgements

The Measurement and Evaluation Center in Reading Education wishes to thank Phi Delta Kappa for granting permission to reproduce this monograph. Gratitude is likewise due to Dr. John J. Horvat, Director of the Laboratory for Educational Development, for the support provided. Mrs. Catherine F. Siffin supervised the editing and production of the manuscript. This book was designed by Dermot McGuinne of the Indiana University Publications Office.

Contents

Preface 3
Introduction 6

Part 1: The Status of Educational Evaluation 7

The Educational Setting 7
The Need for Improved Evaluations 7
Problems in Educational Evaluation 8
The Problem of Definition 9
Problems in Defining Settings to be Served by Educational Evaluations 10
Problems in Defining Decision Types 10
Problems in Designing Educational Evaluations 10
The Problem of Designing Evaluation Systems 12
The Problem of Defining Criteria for Judging Evaluations 12
Summation 12
Footnotes 14

Part 2: A Proposal 15

Premises 15
Evaluation Defined 15
Decision Settings and Evaluation Strategies 18
Small versus Large Change 20
High versus Low Understanding 20
The Decision-Making Settings 21
Types of Decisions 22
Planning Decisions 24
Structuring Decisions 24
Implementing Decisions 25
Recycling Decisions 26
Types of Evaluation 26
Context Evaluation 26
Input Evaluation 27
Process Evaluation 27
Product Evaluation 28
The Structure of Evaluation Design 28
Focusing the Evaluation 28
Collection of Information 30
Organization of Information 30
Analysis of Information 30
Reporting of Information 32
Administration of Evaluation 32
A Total Evaluation System 32
Criteria for Judging Evaluations 33
Internal Validity 33
External Validity 33
Reliability 33
Objectivity 33
Relevance 33
Significance 34
Scope 34
Credibility 34
Timeliness 34
Pervasiveness 34
Efficiency 34
Footnotes 35

Figures
1 The Relation of Evaluation to Decision-Making 16
2 Decision-Making Settings 19
3 Types of Decisions 23
4 Developing Evaluation Designs 29
5 The CIPP Evaluation Model 31

Introduction

This paper was delivered at the Second National Symposium for Professors of Educational Research, sponsored by Phi Delta Kappa at Boulder, Colorado, November 21, 1968. The authors were asked to summarize, synthesize, and update some of their past writings about educational evaluation.

A great deal of confusion and controversy regarding the relationship of evaluation methodology to research methodology exists. The authors of this paper have taken a rather specific position in this controversy, rejecting the proposition that evaluation is equivalent to research, that is, that the same assumptions and methodologies hold for the two fields. The writers assert that professors of educational research are largely to blame for the confusion and ineptness which persist in the field of evaluation. The authors think that many researchers make wrong assumptions about what an evaluation study should accomplish, and that, based on these erroneous assumptions, researchers foist bad advice upon unsuspecting and unsophisticated practitioners. As a consequence, evaluations are usually useless, and practitioners are largely justified in the jaundiced view they typically have taken about evaluation and its utility.

The authors attempt to validate these assertions and to suggest some alternative ways of viewing evaluation. Their aim is to stimulate inquirers and developers to help produce a new methodology which will permit more effective evaluation practice. At the very least, the authors hope to expose some of the more salient issues concerning evaluation and stimulate discussion of them.

The views presented here are drawn from a number of sources, including several of the authors' own papers,[1] the writings of other experts in the field, and especially intensive discussions with a number of colleagues.[2] This paper is divided into two parts. In Part 1 the authors attempt to describe the field of educational evaluation as it exists and to delineate six major problems which must be overcome if evaluation as a science is to be soundly advanced. In Part 2 the authors attempt to respond to these problems. Essentially this response takes the form of a proposed new definition of evaluation and the means through which this definition may be explicated and operationalized. Overall, the paper attempts to point out directions which other research methodologists can follow in advancing the theory and practice of educational evaluation.

The authors wish to emphasize the tentative nature of these formulations, which are still in an early developmental state and are themselves largely unevaluated.

Egon G. Guba
Daniel L. Stufflebeam

Part 1:
The Status of Educational Evaluation

THE EDUCATIONAL SETTING

Education is highly valued as a means for meeting the social, economic, technological, and scientific needs of society as well as the intellectual needs of citizens. To fulfill this complex role educators must deal with a wide range of urgent problems, such as the inequality of opportunity afforded to members of minority groups, riots in the cities, disillusionment of youth, and school dropouts. Education thus has a most difficult charge which requires the initiation of many innovative programs.

To facilitate such educational innovation, society is annually providing billions of dollars through federal, state, and foundation programs to education agencies at all levels. Examples of increased support to education include the Elementary and Secondary Education Act of 1965, the Headstart Program, the Follow-Through Program, The Educational Professions Development Act, and the Experienced Teacher Fellowship Program. Many industries and non-profit research enterprises are also developing education components, and education-industry combines have become common place. Clearly, in addition to new responsibilities, education also has unprecedented opportunities to improve and expand its programs.

Along with the opportunities for change goes a responsibility to evaluate new plans and programs. Evaluation requirements are especially evident in federal assistance programs. Here the law explicitly states that fund recipients will make at least annual evaluation reports.

Such requirements for evaluation are reasonable and long overdue. Funding agencies and the public have a right to know whether their huge expenditures for education are producing the desired effects. And, educators themselves need evaluative information to be sure the changes they induce are in fact improvements.

To justify requirements for evaluation is not equivalent to operationalizing them, however. Educators must respond by establishing evaluation units, defining the roles of staff needed to operate these units, and recruiting and training personnel to fill those roles. They must determine the evaluative questions to be answered, select or construct appropriate instruments, and select samples of the persons who are to respond to the instruments. They must provide means for organizing, analyzing, and reporting evaluative information; and they must define the evaluation schedule, policies, and budget. Last but not least, evaluators must develop working relationships with those who will provide information for the evaluation as well as those who will receive and utilize the information. Clearly, the task of evaluating any educational program is highly complex.

THE NEED FOR IMPROVED EVALUATIONS

How have educators responded to their new evaluation responsibilities? To what extent have they responded at all? And how good have their evaluation studies been?

Without question, educators have made a massive response to requirements for evaluation. The multitude of evaluation reports now available from local schools, state education departments, regional educational laboratories, educational industries, and the like, is a dramatic indication of the significant expenditures of time, effort, and money for the evaluation of educational programs. However, the increased activity alone has not met the need for effective evaluations. While educators have been busy doing evaluations, these evaluations have not provided the information needed to support decision-making related to the programs being evaluated.

Many of the completed evaluation reports contain only impressionistic information. Though such information may be pertinent to the concerns of decision-makers, it usually has lacked the level of credibility required by decision-makers to defend their decisions, and seldom has such information been of material use in arriving at important decisions. A case in point is the first annual report for Title I of The Elementary and Secondary Education Act.[3] This report was highly important as it encompassed thousands of Title I projects throughout the nation. However, it fell far short of being a useful document, for it was almost devoid of hard data. On the other hand, it did contain many anecdotal accounts in which persons who were responsible for conducting Title I activities stated that they felt that their programs had been successful. Many of them speculated as to the reasons for the alleged successes. Though these anecdotes may have touched key issues related to improving the billion dollar per year Title I program, decision-makers in

the Congress, the Office of Education, state education departments, and local school districts could hardly base important decisions on a few "possibly accurate" pieces of testimony.

The situation is not much different in Title III of the Elementary and Secondary Education Act. Title III staff members in the U.S. Office of Education have continuously ranked the quality of Title III projects on a five point scale for each of fifteen criteria.[4] The criterion relating to evaluation has consistently been ranked near the "poor" end of the scale and lower than thirteen of the other criteria—the exception being the criterion related to dissemination. One of the authors of this paper made an analysis of thirty-two Title III projects, and concluded that "it is very dubious whether the results of these evaluations will be of much use to anyone. They are likely to fit well, however, into the conventional school man's stereotype of what evaluation is: something required from on high that takes time and pain to produce but which has very little significance for action."[5]

Unlike the Title I and Title III evaluations referred to before, some evaluations provide for hard data. For example, the evaluation report for New York City's Higher Horizons Program[6] used rigorous research procedures to compare the performance of an experimental group receiving the Higher Horizons Program with the performance of a control group matched to the experimental group on several counts. The basic conclusions contained in this nearly 300 page report were typical of findings for rigorous educational evaluations: "There were no significant differences." In sharp contrast, however, the report also noted that the teachers and principals who had been involved in the program said that it was making differences so significant that the program simply could not be abandoned.

Though the Title I, Title III, and Higher Horizons evaluating differed as to rigor, they were alike in one respect. None of them provided much help to decision-makers for improving the programs being evaluated. While only three examples of the deficiencies in current evaluations have been cited, they are sufficiently important ones to illustrate the point. In too many cases, evaluation reports provide little or no help to decision-makers, and decision-making in and about education must remain an arty endeavor.

PROBLEMS IN EDUCATIONAL EVALUATION

What is the explanation for this situation? Why is it that educators are failing to provide evaluations which are at the same time useful and scientifically respectable? Why is it that evaluations which adhere to classical research methods provide information which is of only limited help in making decisions about programs, and why do the typical "no significant difference" findings in so many of these evaluations contravene the experiences of those who are intimately involved in the programs?

One cannot answer these questions simply on the grounds that evaluation practice lags too far behind evaluation theory, or that there is a lack of effort on the part of educators to evaluate their programs. Further, it is not enough to note that evaluation testimony given by witnesses is not credible, or that typical findings of "no significant differences" are correct because nothing in education ever makes a difference. Rather, the lack of adequate evaluation information probably persists because of **several fundamental impediments** which must be removed before educators can improve their evaluations. These impediments include the lack of trained evaluators and training programs, the lack of appropriate evaluation instruments and procedures, and the lack of adequate evaluation theory. This latter lack is, the authors believe, crucial.

Clearly, the conceptual bases for evaluations are of fundamental importance. If these conceptions are faulty, then the evaluations which are based on them must also be faulty. Thus, it is highly important to identify and examine the efficacy of conceptualizations which underlie current needs for evaluation as well as educators' attempts to meet these needs. It will be useful to divide these conceptualizations into six classes and to consider each one separately. The six classes are:

1. The definition of educational evaluation.
2. The nature of the educational settings within which evaluations must be conducted.
3. The definition of information requirements for educational evaluation.
4. The structure of evaluation designs.
5. The structure of evaluation systems.
6. The definition of criteria for judging evaluations.

The Problem of Definition

Evaluation, like any analytic term, can be defined in many essentially arbitrary ways. The question is not so much that of the "right" way to define evaluation, but how we can recognize the contribution that different definitions can make to our thinking and how to devise a particular definition that suits the purposes and needs in mind.

In its earlier days educational evaluation was largely equated with measurement. This is not surprising in view of the history that preceeded it. Evaluation really came into its own during the twenties following upon the heels of the very successful measurement movement. Abraham Kaplan is fond of what he calls the "Law of the Instrument," viz.,

> Give a small boy a hammer, and he will find that everything he encounters needs pounding.[7]

Thus it was natural that following the successful invention and adoption of standardized tests it was found that everything needed to be tested. The accumulation of scores and the statistical manipulation of those scores to produce that pseudo-standard called norms made it possible to form many judgments; this process came to be called evaluation.

This definition had the advantage of stressing the importance of reliability, validity, and objectivity in collecting and interpreting data; it had great disadvantages in that it ignored the judgmental aspect of evaluation and that it tended to eliminate as unimportant, variables for which instruments were not readily available.

Another definition which received wide currency, and is still the backbone of most evaluative thinking today, is that formulation which regards evaluation as a process for determining the congruence of performance and objectives. All school programs should be guided by behavioral objectives; indeed, it is the essence of program planning to project objectives and the essence of curricular planning to project a series of experiences through which the pupil can achieve the objectives. Similarly it is the essence of evaluation to determine whether the objectives were in fact met.

This definition has certain advantages and disadvantages. Clearly, it is possible in these terms to focus not only on the student but also on the program. If objectives are not met, it is not a foregone conclusion that it is the student who is out of step. Thus feedback is encouraged leading perhaps to diagnosis and remediation of students, but just as possibly, leading to curricular change and refinement.

But at the same time evaluation is pointed by this definition toward **outcomes;** one cannot evaluate until performance has taken place and can therefore be compared to objectives. Thus while the definition implies feedback leading to refinement, such feedback typically cannot occur until the termination of the treatment being evaluated. The promise of the definition is thus often not attained.

Moreover, the matter of criteria remains unresolved. While standards are perhaps implicit in the statement of objectives, the source of the objectives is mystical. It is often asserted that they are "screened" through a philosophy and a psychology but which philosophy and which psychology is unspecified, as is the meaning of the term "screen." Finally, this definition of evaluation places an overwhelming importance on **behavioral** objectives, encouraging the belief that only "ultimate" evaluations which trace everything back to improved student performance are worthy of the name. Today there is insistence on evaluating even national programs like ESEA Title III this way, forgetting that there are other standards that might be applied with greater validity. Thus it would not occur to us to evaluate a carburetor by looking for changes in driver behavior, but in education an analogous process seems to make sense.

A third definition of evaluation tends to equate evaluation with the judgmental process. If the equation of evaluation with measurement can be scored for ignoring the value dimension of evaluation, then surely the equation of evaluation with judgment can be scored for ignoring the processes of arriving at information. Yet this procedure is fairly common, as, for example, in the evaluation processes of accrediting associations such as the North Central Association or the American Association of Colleges for Teacher Education, where the judgment rendered by a visitation team is the evaluation, or in the panel review processes utilized by many funding agencies, including the U.S. Office of Education, for evaluating proposals. While this method has the advantages of quick response and the utilization of the full range of the evaluator's competence,

it obviously leaves much to be desired in terms of objectivity and validity, which are at best moot.

None of these three definitions is thus entirely satisfactory. Each has certain advantages which should be retained, but each also has certain disadvantages which are at best annoying and at worst devastating. Clearly a more defensible formulation is required.

Problems in Defining Settings to be Served by Educational Evaluations

Let us examine the problems involved in providing an adequate focus for educational evaluation studies. Obviously, to evaluate one must know something about the program within which the evaluation is to be conducted. Gaining such knowledge, however, is a difficult task at best. Current needs for educational evaluation have arisen in relation to programs and activities which are new to the field of education. Such activities involve responsibilities newly assigned to educators, new kinds of relationships among different kinds and levels of agencies, and a need for cooperative decision-making about education among a variety of educational and non-educational agencies. It should come as no shock if the evaluation theory which has traditionally been viewed as appropriate for education is found no longer to be adequate to meet the information requirements in new educational settings. Clearly, many of the new programs in education are dramatically different from those of the past, and the new programs themselves differ greatly from each other. Probably different evaluation strategies will be needed for different educational settings. Before these evaluation strategies can be developed, however, the different kinds of educational settings within which evaluations are to be conducted must be conceptualized.

Problems in Defining Decision Types

Even if adequate conceptualizations of the different educational settings to be served by educational evaluation existed, there is insufficient knowledge of the information requirements to be met by educational evaluation. What types of questions must be answered by evaluation studies, and how can they be classified so as to facilitate the development of a generalizable set of evaluation designs? Programs to improve education depend heavily upon a variety of decisions, and a variety of information is needed to make and support those decisions.

Evaluators charged with providing this information must have adequate knowledge about the relevant decision processes and associated information requirements before they can design adequate evaluations. They need to have knowledge about the place, focus, timing, and criticality of decisions to be served. At present no adequate formulation of decision processes and associated information requirements relative to educational programs exists. Nor is there any ongoing program to provide this knowledge. In short, there are no adequate conceptualizations of decisions and associated information requirements or programs to produce them.

Problems in Designing Educational Evaluations

If current conceptions of evaluation are not adequate for evaluating current educational activities, neither can extant designs be adequate.

Recall the kinds of designs educators use to evaluate their programs. If a design is used at all, it typically is an experimental design. The fundamental concern of experimental design is that data which are produced be internally valid, i.e., unequivocal. Several conditions are necessary to meet this criterion. The units to be measured should be randomly assigned to treatment and control conditions. For example, a set of students might be partitioned randomly into two groups—one to receive a new program, the other to receive the school's present offering in the area to be served by the new program. Next, the treatment and control conditions must be applied and held constant throughout the period of the experiment, i.e., they must conform to the initial definitions of these conditions. The new or traditional program conditions could not be modified in process, because in that event one could not tell what was being evaluated. Also, all students in the experiment must receive the same amount of the treatment to which they are assigned; and care must be taken so that students receiving one treatment are not contaminated by the other treatment. If contamination occurred, one could not tell later what had caused what. Therefore, until an experiment is completed, one must resist the temptation to apply the

successful activities of one condition to students receiving a different condition, even if the activities in the latter condition are obviously failing. Finally, an instrument which is valid and reliable for the specified criterion variable must be administered after a certain period of time—usually a complete program cycle—to subjects from both parts of the experiment. Then, if all of the above conditions were met, one could use predetermined statistical procedures and decision rules to determine unequivocally that there were or were not significant differences between the experimental and control groups on the outcome variable of interest.

On the surface, the application of experimental design to evaluation problems seems reasonable, as traditionally both experimental research and evaluation have been used to test hypotheses about the effects of treatments. However, there are four distinct problems with this reasoning.

First, the application of experimental design to evaluation problems conflicts with the principle that evaluation should facilitate the continual improvement of a program. Experimental design prevents rather than promotes changes in the treatment because treatments cannot be altered in process if the data about differences between treatments are to be unequivocal. Thus, the treatment must accommodate the evaluation design rather than vice versa, and the experimental design type of evaluation prevents rather than promotes changes in the treatment. It is probably unrealistic to expect directors of innovative projects to accept conditions necessary for applying experimental design. Obviously, they can't constrain their treatment to its original definition just to ensure internally valid end-of-year evaluative data. Rather, project directors must use whatever evidence they can obtain to refine continually and sometimes to change radically both the design and its implementation. It is thus contended here that conceptions of evaluation are needed which would stimulate rather than stifle dynamic development of programs.

A second flaw in the experimental design type of evaluation is that it is useful for making decisions after a project has run full cycle but almost useless as a device for making decisions during the planning and implementation of a project. The potential confounding the variables must either be controlled or eliminated through randomization if the study results are to have internal validity. However, in the typical educational setting this is nearly impossible to achieve. For example, consider the following quotation from an evaluation report completed by Julian Stanley:

> Even if the program does have considerable cumulative influence on a person's career, this may be slow in appearing and so interactive with other influences that it cannot be discerned clearly by the person himself or by others.
>
> Nevertheless, we must use whatever evidence that can be adduced to determine whether or not such programs are worth repeating and, if so, how they should be modified in order to be more effective. Ideally, in the experimental design sense, we should conduct the program as a controlled experiment, with a well-matched control group that does not attend the institute, and follow up both groups for quite a few years in order to determine how they diverge. If recruiting begins early enough and the applicant group is able enough to provide both groups at a sufficiently high level, this might be done, though the "reactivity" of the disheartened rejectees, the self-fulfilling prophecy of the rejectees, and the inability to control the summer activities of the rejectees might undesirably affect the outcome of the experiment. Merely having on one's record the fact of attending a certain prestigious program, like displaying one's Phi Beta Kappa key, might be a powerful aid. . . . Our chief way of evaluating the success of the program is via reports from staff and participants, particularly the latter.[8]

In the above quotation, Professor Stanley has pointed to many of the reasons why experimental design does not seem well suited to evaluation problems in education. In many innovative programs there clearly are a multitude of confounding factors which simply cannot effectively be controlled.

But the difficulty pointed to here is more complex than one would infer from Stanley's statement. It is not just a matter of being unable, in the real world, to satisfy all the requirements posed by experimental design; it is also a matter of being unwilling to do so. Evaluation is not interested only in determining

the relationship among variables in that best of all possible worlds—the laboratory; it is also concerned with determining what will happen in the worst of all possible worlds. Thus, far from wishing to screen out possible sources of interference, evaluation is actually concerned with inviting interference so that results under the worst possible circumstances can also be assessed.

A fourth problem inherent in the application of conventional experimental design is the possibility that while internal validity may be gained through the control of extraneous variables, such an achievement is accomplished at the expense of external validity. If the extraneous variables are tightly controlled, one can have much confidence in the findings pertaining to how an innovation operates in a controlled environment. However, such findings may not be generalizable to the real world at all because in that world the so-called extraneous variables operate freely. Clearly, it is important to know how educational innovations operate under real world conditions.

The Problem of Designing Evaluation Systems

A fifth problem is that of providing institutional settings in which evaluation can occur as a matter of course. To meet the evaluative needs of educators it is necessary to provide both for continuous, systematic evaluation needs and for unpredictable, ad hoc information needs which emerge in programs of change. Certain routine and predictable information requirements should be provided for systematically just as attendance is taken and achievement data is collected on a regular basis. To handle such information needs de novo each time they occur certainly is inefficient. On the other hand an effective evaluation mechanism should possess the capacity of performing ad hoc studies when they are needed. To meet both of these conditions requires much more knowledge than is presently available about the role of evaluation mechanisms within educational programs. Where should such a unit be housed organizationally? What support is necessary for such a unit? What data should be collected routinely? What evaluative services should be performed for other role functionaries within the educational program? What policies and guidelines should govern the operations of the unit? These and many other related questions should be answered if educational agencies are to install and maintain the types of evaluation units they need.

The Problem of Defining Criteria for Judging Evaluations

Finally, attention must be given to the matter of defining criteria for judging the worth of evaluations. If inappropriate or insufficient criteria are applied for this purpose, serious trouble will result. The result may well be faulty designs and useless reports. If, for example, an evaluation design is selected solely upon the basis of reliability and validity, valid and reliable information might be produced at a time when it is too late to be of any use in an action program. Consider the following excerpt from testimony pertaining to Title I evaluations given before a Congressional committee by a citizens' group in New York City:

> We ask for amendments to render the required evaluations of Title I projects meaningful. The Act states that evaluations must be made, not that they be utilized in future planning. In New York City this year, projects were recycled before last year's evaluations were submitted. To be made more useful, evaluations should have built into them alternatives and the recommendations of the evaluator. What is now an expensive exercise should be made a function to provide service to local school boards having the responsibility for making policy based on experience. American business would not survive if its consultants did not supply management with alternatives after reviewing the efficacy of of programs.[9]

Here, the major concern seems to be that reports yielded by current evaluation programs are neither sufficiently specific nor timely to influence educational programs. Obviously, evaluations which do not at least meet these two criteria are of little use.

SUMMATION

This then concludes a review of the current needs and problems in educational evaluation. The authors have noted that completed evaluations have been far from adequate and have

asserted that the fundamental problem is a lack of adequate conceptualizations regarding the nature of educational evaluation in the context of the emergent programs of educational change. In this regard, six theoretical problems were discussed which the authors believe must be solved before meaningful evaluation methodology can be developed. These problems were:

1. Inadequacies of present definitions of educational evaluation.
2. A lack of understanding of the different educational settings within which evaluation must be conducted.
3. A lack of understanding of generalizable information requirements which educational evaluation studies must meet.
4. The lack of a valid structure for the generalizable parts of evaluation design.
5. The lack of concepts needed to organize and operate evaluation systems.
6. The lack of an appropriate set of criteria for judging the worth of evaluation strategies, designs, instruments, reports, etc.

In the next part of the paper a response is made to each of these problems by proposing some new conceptualizations to undergird the evaluations which the authors believe are needed in programs of educational change.

Footnotes

[1]See especially Daniel L. Stufflebeam, "Evaluation as Enlightenment for Decision-Making," an address delivered at the Working Conference on Assessment Theory, Sponsored by ASCD Sarasota, Florida, January 1968; and Egon G. Guba, "Evaluation and Change in Education," a paper prepared for the Elk Grove Training and Development Center, Indiana University, NISEC, May 1968.

[2]The authors' ideas have profited from the interchange afforded by the Phi Delta Kappa National Study Commission on Evaluation, whose membership consists, in addition to the present authors, of Walter J. Foley, William J. Gephart, Robert L. Hammond, Howard O. Merriman, and Malcolm Provus. The writers have also benefitted greatly from discussions with John Horvat, David L. Clark, and Sidney Eboch.

[3]Public Law 89-10: The Elementary and Secondary Education Act of 1965, Title I.

[4]These criteria are listed on pp. 70-71 of the current Title III guidelines. *(A Manual for Projects Applicants and Grantees,* Washington, D.C.: U.S. Office of Education, 1967).

[5]Egon G. Guba. *Evaluation and the Process of Change,* Notes and Working Papers Concerning the Administration of Programs authorized under Title III of Public Law 89-10, The Elementary and Secondary Education Act of 1965 as amended by Public Law 89-750, April 1967, p. 312.

[6]Wayne J. Wrightstone, et al. *Evaluation of the Higher Horizons Program for Underprivileged Children,* Cooperative Research Project No. 1124, Bureau of Educational Research, Board of Education of the City of New York.

[7]Abraham Kaplan. *The Conduct of Inquiry,* San Francisco: Chandler Publishing Company, 1964, p. 28.

[8]Julian Stanley. *Benefits of Research Design; A Pilot Study,* Final Report, Project No. X-005, Grant OE5-10-272, U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Research, August 1966.

[9]*Citizens' Committee for Children of New York, Inc. Newsletter,* Statement of Mrs. Nathan W. Levin, Chairman of the Educational Services Section, before the Sub-Committee on the Elementary and Secondary Education Act of the Education and Labor Committee of the House of Representatives, March 18, 1967.

Part 2:

A Proposal

To respond to the problems identified in Part 1, this second part of the paper is divided into eight major sections. In the first section the premises are presented upon which subsequent conceptualizations are based. In the second section a general definition of educational evaluation is proposed. Section three contains conceptualizations of different educational settings and of corresponding evaluation strategies required to deal with them. Section four presents conceptualizations of four types of educational decisions, and section five proposes four different evaluation designs appropriate to them. Section six is an effort to outline the structure of evaluation design. In section seven the authors attempt to synthesize their conceptualizations of evaluation strategies, types, and design steps into a single model for an evaluation system which can meet continuous information requirements of an ongoing educational program while still retaining responsiveness to emergent, idiosyncratic evaluation needs. Finally, in section eight the writers suggest and define criteria which can be employed to judge the worth of developed evaluation systems, study designs, and reports.

PREMISES

The general logic of the proposed model is shown in Figure 1. Program operations or activities are evaluated to influence decisions which influence program operations which are in turn evaluated, ad infinitum. Figure 1 also indicates that the evaluation process includes five steps: (1) focusing the evaluation to identify the questions to be answered and the criteria to be employed in answering them, (2) collecting information, (3) organizing information, (4) analyzing information, and (5) reporting information.

Implicit in this logic are five premises which form the basis for the proposed evaluation model:

1. The purpose of evaluation is to provide information for decision-making; to evaluate, therefore, it is necessary to know what decisions are to be served.
2. Different evaluation strategies are required depending upon the nature of different decision-making settings to be served; thus, a valid model for educational evaluation should be grounded in sound conceptualizations of the different decision-making settings to be served.
3. Within any decision setting, different types of decisions require different types of evaluation designs; therefore, a generalizable and efficient evaluation model should be based upon a parsimonious conceptualization of the types of decisions and evaluation designs which are generalizable to all educational decision settings.
4. While the content of different evaluation designs varies, a single set of generalizable steps can be followed in the design of any sound evaluation.
5. Because evaluation studies should answer questions posed by decision-makers, designs for such studies should satisfy criteria both of scientific adequacy and of practical utility.

EVALUATION DEFINED

Given these premises, evaluation is defined as follows:

> **Evaluation is the process of obtaining and providing useful information for making educational decisions.**

This statement contains six key terms. Each of these terms will be defined at this point as each one has significant implications for the processes and techniques of evaluation.

Process is defined as a particular and continuing activity subsuming many methods and involving a number of steps or operations.

Particular attention should be paid to the fact that the evaluation process is conceived as continuing; in particular, it is not conceived as terminal or as having a discrete beginning and ending. It will be seen that evaluation activities tend to be (a) sequential, i.e., with each activity forming a logical base for the next, and (b) iterative, i.e., recurrent or cyclical. Evaluation is also conceived as multifaceted, involving many different methods or techniques. This dynamic, complex conception of evaluation is in sharp contrast to the relatively static, terminal, single-phase conception of evaluation that is current.

Obtaining is defined as making available through such processes as collecting, organizing, analyzing, and reporting, and through such formal means as measurement, data processing, and statistics.

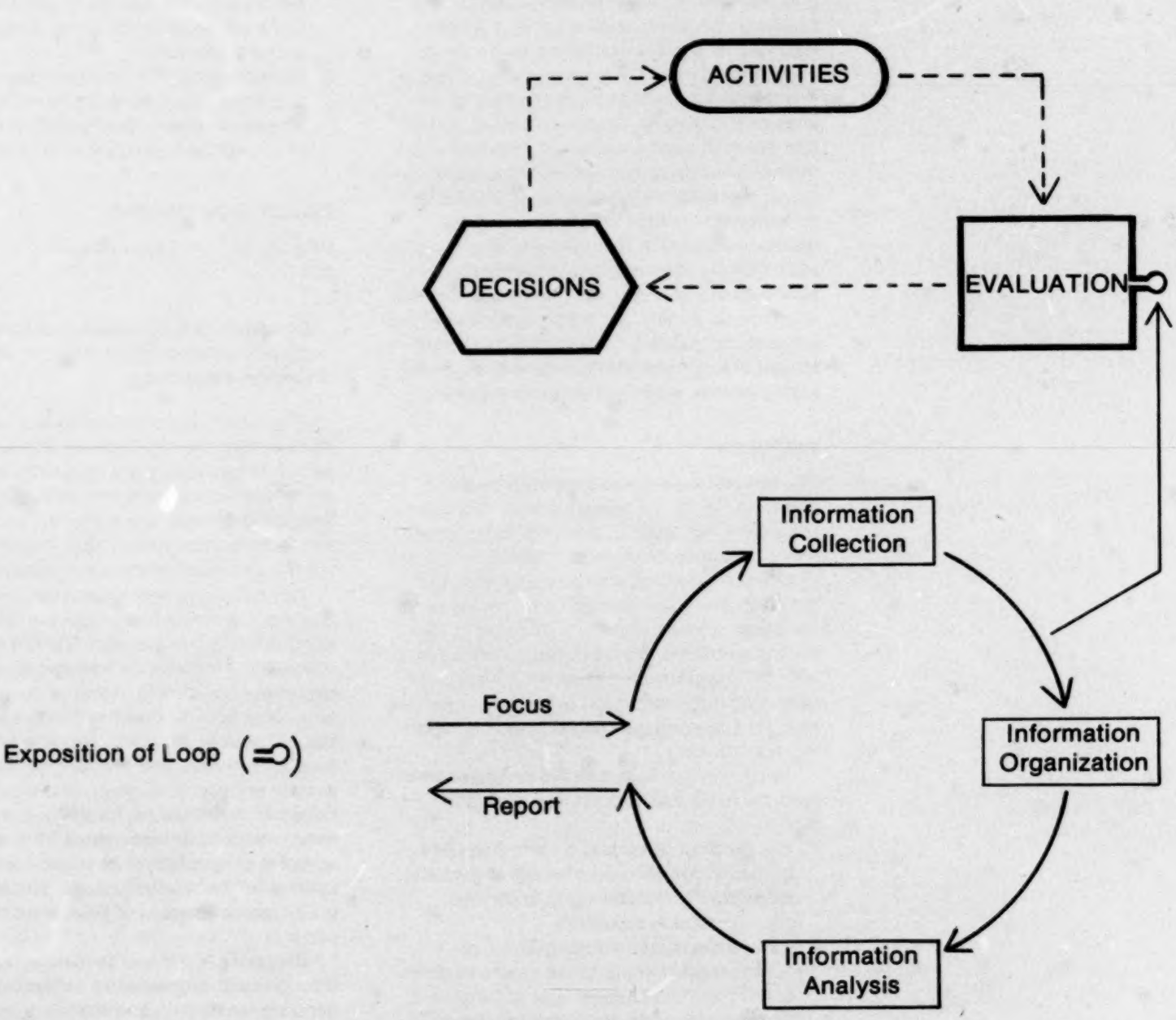

Figure 1: The Relation of Evaluation to Decision-Making

Providing is defined as fitting together into systems or subsystems that best serve the needs or purposes of the evaluation. The evaluator as provider is concerned primarily (but not exclusively) with meeting the ad hoc criteria posed by his client, whatever those may be, e.g., cost, staff availability, political viability, and the like. To provide implies familiarity with conventional techniques of information reporting and transmission, as well as a concern for developing new methods of client criterion identification and the adaptation of information thereto. The evaluator who acts as provider functions primarily as an interactor or interface, a fact which is perhaps the chief basis for the evaluator's claim to a professional role, in the sense of a privileged relation to a client. It is his function to help the client identify his needs, to formulate his criteria, and then to order and highlight the evaluative data into reports that best serve those needs in the framework of the evolved criteria.

Useful is defined as appropriate to established criteria agreed upon by the evaluator and the user or client. In determining utility the evaluator leans heavily upon certain practical criteria such as timeliness and relevance. Evaluation designs must be shaped by considerations of utility rather than being simply helter-skelter collections of easily available or easily measured variables.

Information is defined as descriptive or interpretive data about entities (tangible or intangible) and their relationships. Webster defines information as "knowledge acquired in any manner; facts; data; learning; lore."[10] This definition is useful in reminding readers that evaluation is concerned not only with scientific findings but also with information resulting from precedent or experience. The definition also serves to remind that information can be derived in a variety of ways. Further, the phenomenology which the information purports to describe or relate need not always be measureable in the rigorous sense; so-called intangibles are also eligible for inclusion when required. If conventional methods of obtaining information do not permit measurement of intangibles, it is time to extend the methodology rather than to exclude the difficult variables.

Educational Decisions are defined as choices among alternatives.

The term **decision** is the key term in the entire definition of evaluation. As will be seen, the derivation of the decision situations to be served by an evaluation serves as the touchstone for the design of all evaluation steps and as the ultimate criterion for inclusion or exclusion of any information or technique which might be proposed. The evaluator as provider is concerned with defining, together with the client, the necessary decision questions and the alternatives which exist within each decision situation. The evaluator as obtainer is concerned with the collection, organization, analysis, and reporting of information that illuminates each alternative, weighing each one in terms of its utility as a decision alternative, applying information to the selection of an alternative, testing the selected alternative for utility, and suggesting ways in which the alternative might be improved or further refined or abandoned in favor of some other alternative.

So much for formal definitions. The reader may be tempted to suggest that the proposed definition of evaluation based on the decision-making process is so different from current formulations that it ought to be given a different name than evaluation. He may well feel that putting a new label on this process will prevent many misunderstandings. After all, everyone now conjures up certain mental images, and perhaps certain attitudinal responses, when the term evaluation is mentioned; would it not be wise to avoid all the misinterpretations and false imputations that might result from confusion in the reader's mind between what is meant here and what he has always understood by the term?

No doubt some clarification might occur through this device, but a great many confusions much more detrimental in the long run would probably result. First of all, if the process is not *called* evaluation, it will not be associated in the reader's mind with the requirements for evaluation that are being posed all around him. He will not understand that these formulations are responsive, for example, to the requirement that he evolve a mechanism to evaluate his Title III project. Thus the reader may mistake the arena in which these formulations have application.

Second, if some other name were to be used, the reader might come to believe that the process discussed here is *bigger than* evaluation; evaluation might be seen as only a part of

a more complex process designed to aid decision-makers. Readers will come to see that evaluation, as that term is ordinarily used, looks very much like what is called product evaluation later in this paper. Product evaluation serves a particular kind of decision need (recycling decisions), to be sure, but there are other kinds of decisions as well. It would be unfortunate indeed if evaluation were allowed to carry only this restricted meaning.

Third, and perhaps most important, the process described as evaluation here comes much closer to the root meaning of the term, to evaluate, than does the process which currently masquerades under the name; one might argue that if a name were to be changed it ought to be that of present practice. Values come most meaningfully into play when there are choices to be made, and the making of choices is the essential act of decision-making. What is proposed here is that the entire act of evaluation center on the criteria to be invoked in making decisions. It is through exposing such criteria that guidance is obtained about the kinds of information which should be collected, how such information should be analyzed, and how it should be reported. The term evaluation seems to be particularly suited to the process as described here, as that process makes such distinctive use of value concepts.

Finally, it may be asserted that the proposed definition is not as foreign to contemporary thought as it might at first appear. Similar definitions have been proposed elsewhere. Cronbach, for example, offers this formulation:

> To draw attention to its full range of functions, we may define evaluation broadly as the collection and use of information to make decisions about an educational program.[11]

Thus, the use of the term evaluation in this new sense seems more than justified on the grounds that the arena for evaluation continues to enjoy unequivocal demarcation; that the term is maintained in its broadest meaning; that the root of the term, value, receives maximum prominence; and that the proposed usage already has some currency.

DECISION SETTINGS AND EVALUATION STRATEGIES

Given this new definition of evaluation with its emphasis on decision-making, it is necessary to describe educational decision-making to provide a basis for conceptualizing a relevant methodology for evaluation. The first step will be to define different decision-making settings which, the authors believe, require different evaluation strategies. In this connection, formulations are based heavily on the work of Braybrooke and Lindblom in the area of public policy.[12]

Figure 2 summarizes a conceptualization of four generally different decision-making settings in education, differentiated through the intersection of two continua: "small" versus "large" educational change, and "high" versus "low" understanding to support change. The utility of these two continua arises directly as a consequence of the authors' definition of evaluation. First, it is clear that the rigor and extensiveness of an evaluation is likely to be determined by the importance of the decision which is to be serviced; this importance in turn is gaged by the significance of the change to be brought about through the execution of the decision. Decisions with unimportant consequences clearly would not demand the expense and thoroughness in an evaluation study that would be required by decisions which will have serious consequences. Second, as an evaluator goes about determining what information he should obtain and provide, he must have in mind the information that is already available and the ability of his client to use it in its present form. Evaluations must be more extensive when there is little information available (or when the client cannot use available information in its present form).

Given this brief rationale to justify the introduction of these two continua—amount of change and degree of understanding—let us look at each in greater detail. They are combined to produce the four different decision-making settings of Figure 2, which require four different evaluation strategies to service them.

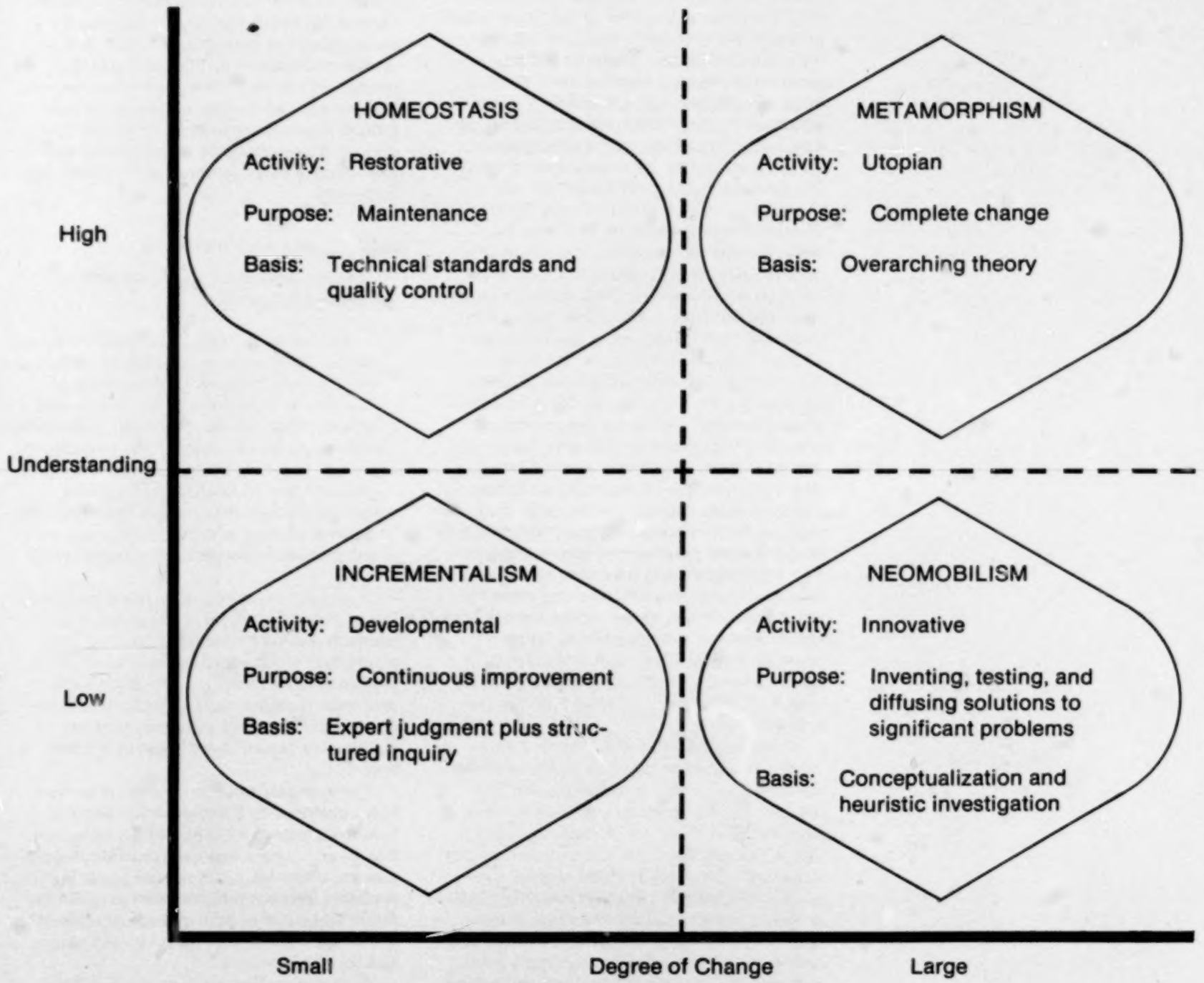

Figure 2: Decision-Making Settings

Small versus Large Change

The authors' basic rule for distinguishing between "small" and "large" change is that the more controversial a change the larger and more important it is. The reader can see how this rule works by focusing on the issues posed in several disagreements in our society about the efficacy of present programs and recommended changes. Some of these come to mind immediately, e.g., a federally-controlled education system, federal aid to public education, busing of children for equalizing educational opportunities, decentralization of large city school districts, automation and "dehumanizing" of education, learning aided by chemotherapy, National Assessment, etc. While it would be very difficult to obtain concensus positions with regard to any of these issues, it would not be difficult to obtain agreement that each one is important. Changes in these and similar areas could result in major restructuring within education, but more significantly such changes could potentially produce results in the lives of individuals and in society at large which are at great variance with the results being produced by the present educational offerings. Many persons would view these changes as potentially so damaging that they would counsel—or demand—that such changes be introduced cautiously and gradually or that they be installed only after sufficient tryout information had been obtained to allow reliable and valid predictions of the effects of the changes. These issues and others like them, signify what the authors mean by "large" changes. They involve major restructuring of education and potentially can have significant impact on variables considered important by society.

At the opposite end of our continuum, "small" changes are identified as unimportant variables. Examples include employing new teachers to fill present vacancies, purchasing a new edition of the textbook series currently in use, increasing the amount of attention devoted to fractions in the current mathematics program, adding wrestling to the present athletic program, replacing all the blackboards with green ones, replacing the present achievement testing series with a new one, adding a public relations official to the school staff, decreasing the average class size, blacktopping the school playground, etc. Changes such as these would stimulate relatively little disagreement. Yet, taxpayers and educators alike would agree that decisions to effect even small changes such as these should not be made blindly. They should be supported by information which will increase the likelihood of choosing the most efficient and effective alternatives, whether the choice concerns the hiring of a teacher or the selection of textbooks. Small changes, then, are changes within education which potentially will have no significant impact on variables considered important by society. Small changes also are characterized by being serial in nature; they result in small, stepwise shifts rather than large ones.

High versus Low Understanding

The second major continuum proposed by Braybrooke and Lindblom is:

> . . . the degree to which the decision-makers can be supposed to understand all the features of the problem with which they are faced. Near one extreme, information is generally lacking; values (goals, objectives, constraints, side conditions) are neither well understood nor well reconciled, and intellectual capacity generally falls far short of grasping and thinking through the problem. Near the other extreme, all aspects of the problem are quite well grasped in the decision-maker's mind.[13]

It is important to note from this quotation that high understanding is composed of two elements: relevant information and the decision-maker's intellectual capacity to utilize that information in the solution of practical problems. Both elements are important in decision-making and the evaluator must be equipped to cope with deficiencies in either one.

The evaluator's role as obtainer is particularly influenced by the information element. Two requirements must be met if information is to be adequate: a validated theoretical structure and adequate, practical data about the particular decision problem must exist. To the extent that either or both of these requirements is not met, the evaluator must strive to obtain additional information.

As provider of information, the evaluator must be particularly concerned with the ability of the decision-maker to understand both the

theoretical and practical information which is available to him. Even if there is adequate information available, it can have little positive influence on decision-making if it is in a form that the decision-maker cannot understand. Therefore, the evaluator has a very critical role in fitting together, assessing, and translating available information which has potential relevance to the needs of the decision-maker.

It is thus clear that the design of evaluation should be grounded in knowledge about the amount and importance of change to be effected and the amount and quality of understanding which is available to support decision-making to effect the change. Only in this way can the evaluator be confident that his study will be useful.

The Decision-Making Settings

Figure 2 is suggested as an aid in understanding the general classes of decision-making settings within which evaluation studies must be conducted. The "small versus large change" and the "high versus low understanding" continua have been combined to yield four classes of educational decision-making settings: decisions to effect large changes supported by a high level of relevant understanding (the upper right cell of Figure 2: Metamorphism); decisions to effect small changes supported by a high level of relevant understanding (the upper left cell of Figure 2: Homeostasis); decisions to effect small changes supported by a low level of relevant understanding (the lower left cell of Figure 2: Incrementalism); and decisions to effect large changes supported by a low level of relevant understanding (the lower right cell of Figure 2: Neomobilism).

Metamorphic decision-making denotes utopian activity aimed at producing complete changes in an educational system. Its guiding basis is an overarching theory which is necessary and sufficient to every detail of the proposed change, and which is completely understood in all its ramifications by the decision-maker. The decision-maker, moreover, must be capable of considering all relevant variables and collecting, analyzing, and synthesizing performance data about these variables as the change is being managed.

The probability favoring this kind of change in any educational institution is indeed slim. Rarely exists the utopian situation in which adequate theory and information systems to support the application of the theory are present along with decision-makers who can assimilate and use the theory and the necessary information as a rationale to effect revolutionary changes. To the extent that such contexts might exist, evaluation strategies needed to support them could be mainly of the total information management system type. The adequate supply of knowledge which already existed would be organized and stored for rapid retrieval whenever the decision-maker might call for it.

Obviously, such utopian educational decision-making settings are mainly theoretical. Therefore, they will not be dealt with further here. However, theoretical identification of such a setting serves a function as educationists and especially critics of education are prone to act as if such change settings do in fact exist. That is, many assume that adequate theory and information are available for effecting whatever utopian changes might be desired and that decision-makers can obtain, understand, and use this information appropriately. If professional educators did not assume this, they certainly would take more pains to collect information to support the large changes they do attempt.

Homeostatic decision-making denotes restorative activity aimed at maintaining the normal balance in an educational system and guided by technical standards and a routine, cyclical data collection system. Of the four types being considered, settings of this type are the most prevalent in education. The major function of educational administration and supervision is to maintain the normal balance in the program, that is, to control the activity and to make adjustments as required to adhere to the specifications established for the program. Staff assignments, scheduling of students, and establishment of bus routes illustrate this type of decision-making.

Homeostatic decision-making settings require evaluation systems characterized by technical standards and quality control data collection systems. The most prevalent forms of routine data collected for homeostatic decisions include achievement data, attendance records, pupil-personnel data, staff records, and community census data. Most schools have adequate quality control evaluation systems to service their homeostatic decision needs. Further,

the changes effected by these decisions are small and remedial. All in all, no major breakthroughs in evaluation theory are needed to service such minor adjustments which are already based on adequate supplies of information. Therefore this setting will not be considered in further detail.

Incremental decision-making denotes developmental activity having as its purpose continuous improvement in a program. Such activity usually is supported by expert judgment and structured inquiry into the efficacy of the present program and the recommended changes. Decision-making in this quadrant differs from homeostatic decision-making in two respects. First, incremental decisions are intended to shift the program to a new normal balance based upon small, serial improvements, while homeostatic decisions are intended to correct the program and change it back to its normal balance. Second, while homeostatic decisions are supported by technical standards and a continuing supply of routinely collected information, evaluations for incremental change are usually ad hoc and supported by little extant knowledge. Special studies, the employment of expert consultants, and the formation of special committees characterize most efforts to introduce incremental change.

Incremental decision-making is very prevalent in education. Many so-called educational innovations are of the incremental type. They are attempts to make improvements in the present program without risking a major failure. Though there is little information to support such changes, the adjustments are sufficiently small that corrective adjustments can be made as problems are detected. As might be expected, such changes are based on trial and error and are iterative and serial in nature. Also, such changes often require allocations of special resources. Title I of the Elementary and Secondary Education Act fosters much incremental change. "Congruence evaluation" systems are needed to support incremental change. Basically, such evaluation programs focus on the congruence between intended and actual increments of program change.

Neomobilistic decision-making denotes innovative activity for inventing, testing, and diffusing new solutions to significant problems. Such change is supported by little theory or extant knowledge; yet, the change is large, often because of great opportunities such as those being produced by the knowledge explosion, or because of critical conditions such as riots in inner cities. Evaluation systems to support neomobilistic decision-making usually are ad hoc, non-rigorous types of investigations. Often, these studies are exploratory and heuristic in nature.

Neomobilistic decision-making is becoming more prevalent in education. Critics of education who advocate higher rates of change, the explosive conditions in our cities, and the knowledge explosion, are all aspects which have served to motivate this kind of change. Title III projects and educational policy research centers engaged in long-range educational planning are illustrations of expenditures of risk capital to stimulate educators to create and to try out new ideas. To support this kind of change, "contingency evaluation" systems are needed. Such systems should be heuristic. They should explore opportunities and possibilities. And, they should stimulate inventions of new solutions to critical educational problems.

TYPES OF DECISIONS

Knowledge of the four decision-making settings is a necessary but not sufficient condition for formulating an evaluation model capable of serving decision-making. For within each decision-making setting one could identify literally thousands of specific educational decisions, all of which might differ from each other in certain respects. Unless ways can be found to group these individual decisions, it will be necessary to contrive a different design for every conceivable decision. Then the notion of generalizable evaluation designs would be meaningless, and the development of evaluation designs would always be ad hoc. Thus, the need is to devise a typology or taxonomy of decisions whose categories are exhaustive of all possible educational decisions while also being mutually exclusive. Under these circumstances, generalizable evaluation designs to fit all decision types within similar categories become feasible.

Figure 3 presents the conceptual base from which the typology proposed is generated. The authors postulate first that decisions should be classified as a function of whether they pertain to ends or means; this fact is depicted by the row headings of Figure 3. The column headings

	Intended	Actual
Ends	PLANNING DECISIONS to determine objectives	RECYCLING DECISIONS to judge and react to attainments
Means	STRUCTURING DECISIONS to design procedures	IMPLEMENTING DECISIONS to utilize, control and refine procedures

Figure 3: Types of Decisions

portray the second dimension which enters into the typology: relevance of the decision to intentions or actualities. Thus, all educational decisions may be exhaustively and unambiguously classified as pertaining to **intended ends—goals, intended means—procedural designs, actual means—procedures in use,** or **actual ends—attainments.** As will be noted, this schema allows the identification of four types of educational decisions which are respectively serviced by four special types of evaluation: **Planning decisions** to determine objectives; **structuring decisions** to design procedures;**implementing decisions** to utilize, control, and refine procedures; and **recycling decisions** to judge and react to attainments. Each type of decision is considered at length.

Planning Decisions

Planning decisions specify **major changes that are needed in a program.** The need for planning decisions arises from awareness of a lack of agreement between what the program was intended to be and what it actually is, or awareness of a lack of agreement between what the program could become and what it is likely to become. In either case, decisions could be made to change or not to change either intentions or actualities, pertaining either to means or ends. Any such decision to introduce change would result in the establishment of program objectives. In this paper objectives which pertain to changes in either the intended or actual means will be referred to as **instrumental objectives** and objectives which pertain to changes in either the intended or actual ends will be referred to as **consequential objectives.** (Behavioral objectives are one example of the possible types of consequential objectives.)

Planning decisions are illustrated by the following questions: Should program goals be changed? Should the present mission be sustained or changed? What are the top priority needs that the program should serve? What are the characteristics of the problems which must be solved in meeting the top priority needs served by the program? What behaviors should the students exhibit following their participation in the program?

As may be inferred from these examples, the authority for planning decisions usually, but not always, resides with policy groups. Thus, the authors come close to equating planning decisions with policy decisions. Role functionaries who make such decisions include boards of education, school superintendents, state superintendents of public instruction, department chairmen, boards of regents, deans, college presidents, regional educational laboratory directors, directors of research and development centers, the commissioner of education, and the like. Obviously, teachers also make planning decisions with regard to behavioral outcomes.

The formulation of planning decisions has consequences that are both internal and external to the program of interest. Consequences that are internal to the program would usually take the form of directives sent from policy figures to subordinates. Such directives would give notice of new objectives and likely would specify modifications in program functions in order that the objectives be achieved. Consequences that are external to the program of interest would usually be in the form of proposals to funding agencies or other external groups which might have the capacity to aid or constrain the program. Such proposals would likely seek funds, sanction, and/or endorsement. Clearly, planning decisions are of fundamental importance to any program, and appropriate evaluation mechanisms should be maintained to provide information for the formulation of new objectives or the modification of existing ones.

Structuring Decisions

Structuring decisions specify the **means to achieve the ends** which have been established as a result of planning decisions. Specification of means must consider variables such as method, content, organization, personnel, schedule, facilities, and budget. Decisions about such variables arise from three sources: awareness of planning decisions which specify what the program is to achieve, awareness that there are alternative means available to achieve the specified outcomes, and awareness of the relative strengths and weaknesses of the available procedural alternatives. Given these three conditions, an action plan to achieve the desired objectives can be structured.

It is noteworthy that structuring decisions can result in the modification of the established objectives. For while objectives are initially based on needs, problems, or opportunities,

they may be and frequently are modified because of realistic limitations on available means to insure their achievement. For example, the aim is to eliminate the possibility of further assassinations of American presidents. While most citizens likely would support this objective in the abstract, many of them clearly would reject the notion of achieving the goal at all costs. The National Rifle Association members and weekend hunters certainly will not succumb easily to a law which would prohibit them from owning fire arms, and certainly no President will submit to solitary confinement. Yet these or similar means are theoretically available to achieve this objective. Instead of supporting these radical means, gun enthusiasts might readily support a law which would restrict known criminals from owning weapons, and future Presidents might yield to a regulation that they participate in parades only in bullet-proof, enclosed vehicles. While such means would not eliminate the threat of assassination, they would materially reduce the probability of such an occurrence. As seen in this example, the relative acceptability of available alternative means and their associated possible outcomes can serve effectively to modify specified outcomes. Ends determine means, and vice versa.

An action plan based upon structuring decisions is a comprehensive statement of outcomes to be achieved, work to be performed, and resources and time to be used. The specified outcomes are those given by the planning decisions, possibly as modified by structuring decisions in the selection of means. The decisions pertaining to work, resources, and time take the form of PERT networks, job descriptions, line-staff organizational plans, procedural specifications, process and product evaluation designs, and program budgets. Collectively, such decisions provide the operating guidelines needed to respond effectively to planning or policy decisions.

Most structuring decisions are made not by policy makers but rather by operations managers. Such managers include project directors, school principals, activity supervisors, and area coordinators. The usual function of policy administrators in structuring decision processes is to confirm that the structuring decisions of their subordinates are consistent with the established policy structure.

Consequences of structuring decisions generally include actions to operationalize a procedure. Budget is allocated. Staff are recruited and oriented to the intents of the activity. Needed materials and facilities are obtained and prepared. Management and clerical procedures are developed, and responsibilities are assigned. Finally, as any operations manager will readily attest, a major time consuming consequence of structuring decisions are seemingly endless meetings and presentations designed to orient staff and create interest and goodwill among the activity's various publics.

Implementing Decisions

Implementing decisions are those involved in **carrying through the action plan.** These decisions arise from two sources: knowledge of the procedural specifications, and continuing knowledge of the relationship between procedural specifications and the actual procedures. These two kinds of information aid in process control.

Implementing decisions involve many choices regarding changes of ongoing procedures. Questions illustrating this type of decision include: Should the staff be retrained? Should new procedures be instituted? Should additional resources be sought? Should responsibilities be reassigned to staff? Should the schedule be modified? Should the public relations activities be changed? Obviously, the making and execution of implementing decisions comprise much of the day-to-day responsibilities of operating any program.

Authority for the implementing decisions is vested largely in operations managers and their designated representatives. Largely, these are the same role functionaries that were identified above for the structuring decisions. Additionally, those responsible to the operations manager such as teachers and counselors hold delegated powers as a part of their roles to make certain implementing decisions.

Implementing decisions also have varied consequences. A role functionary performs his work differently. Inservice training sessions are conducted. Staff obtain a better understanding of their individual and collective roles. Specialists external to the program are consulted. Newspapers publicize certain aspects of the program. New personnel are added to the staff. Personnel work overtime. The PERT schedule is updated. New materials are obtained. Facilities are adapted to emergent program needs.

Although many of these consequences seem routine, it is clear that their cumulative effect can largely determine the success of a program. Therefore, operations managers need daily access to information which can shape their implementing decisions.

Recycling Decisions

Recycling decisions are the fourth and final type of decisions in our classification schema of educational decisions. These decisions are those **used in determining the relation of attainments to objectives** and in determining whether to continue, terminate, evolve, or drastically modify an activity. The essential type of awareness precipitating these decisions is knowledge of the nature and timing of specified attainments.

Basically, recycling decisions involve product control choices. Such decisions are usually thought to occur after a complete cycle of an activity. But this is a limited view of recycling decisions. More appropriately conceived, they occur throughout an activity as quality or product control devices. Therefore it is emphasized that recycling decisions are concerned with attainments at any point in a program as opposed solely to outcomes following a full cycle of a program. Whereas implementing decisions focus on the extent to which means are operant as intended, recycling decisions focus on the extent to which desired ends are being and/or have been attained.

Many questions illustrative of what is meant by recycling decisions can be posed. Are the students' needs being met? Are problems being solved as intended? Is the program failing? Was the outcome worth the investment? Has there been a significant gain in pupil achievement? Has the program benefited by using the opportunity that was presented? Has sufficient progress been achieved to warrant continuation of the program? Is the new program succeeding? Were the results from Program A better than those from Program B? Was the procedure effective? Has the program resulted in improved teacher competence? Have school-community relations been improved? Have students improved their self-concepts? Questions such as these often must be answered when operations managers are attempting to justify new funding requests. Continuing to fund expensive procedures without answering such questions understandably is often frowned on by responsible fiscal agents.

Authority for recycling decisions usually resides with the operations manager during the implementation of an activity cycle and with the responsible fiscal agent at the conclusion of an activity cycle. While the operations manager can make certain decisions about outcomes which might have policy or fiscal implications, he usually has very limited authority to make recycling decisions which would result in major policy or fiscal changes. Therefore, the policy maker is a key figure to be involved in recycling decisions.

Recycling decisions have very tangible consequences. Program activities may be continued at the same level of funding under the same product specifications; they may be drastically changed; or they may be discontinued. New funding proposals often are written as a result of recycling decisions. Present staff may be reassigned or discharged. Attempts may be initiated to diffuse or install the tested product into a broader context. The activity cycle that produced the product may be debugged and recycled.

TYPES OF EVALUATION

Corresponding to each of these four decision types are four types of evaluation, which might be thought of as four generalizable evaluation designs; the four types are given the names **context, input, process, and product.** It might be noted that the initial letters of these four terms form the acronym CIPP (pronounced sip) which is often used as a general name for the formulations which are propounded here. **Context** evaluation services **planning decisions,** input evaluation services **structuring decisions, process** evaluation services **implementing decisions, product** evaluation services **recycling decisions.**

Context Evaluation

The major objective of context evaluation is to define the environment in which change is to occur, to depict unmet needs, and to identify the problems that result in needs not being met. For example, the environment might be defined as the inner city elementary schools of

a certain metropolitan area. A context evaluation might reveal that children were not learning to read at the level expected of them, and it might further indicate that a particular problem, or problems, was the cause of this failure, e.g., instruction might be inadequate, materials might not be appropriate, a language barrier might exist, there might be a high rate of absenteeism, and the like. Thus the children's need to learn to read was being thwarted by certain particular problems. Environment, needs, and problems would all be involved in the context evaluation.

The method of context evaluation begins with a conceptual analysis to identify and define the limits of the domain to be served as well as its major subparts. Next, empirical analyses are performed, using techniques such as sample survey, demography, and standardized testing. The purpose of this part of context evaluation is to identify the discrepancies among intended and actual situations for each of the subparts of the domain of interest and thereby to identify needs. Finally, context evaluation involves both empirical and conceptual analyses, as well as appeal to theory and authoritative opinion, to aid judgements regarding the basic problems underlying each need.

Input Evaluation

The major objective of input evaluation is to determine how to utilize resources to meet program goals. This objective is accomplished by identifying and assessing relevant capabilities of the proposing agency, strategies which may be appropriate for meeting program goals, and designs which may be appropriate to utilize a selected strategy. The end product of input evaluation is an analysis of alternative procedural designs in terms of potential costs and benefits. Specifically, alternative designs are assessed in terms of their staffing time, and budget requirements; their potential procedural barriers; the consequences of not overcoming these barriers; the possibilities and costs of overcoming them; relevance of the designs to program objectives; and overall potential of the design to meet program goals. Essentially, input evaluation provides information for deciding whether outside assistance should be sought for meeting goals and objectives; what strategy should be employed, e.g., the adoption of available solutions or the development of new ones; and what design or procedural plan should be employed for implementing the selected strategy.

Methods for input evaluation are lacking in education. The prevalent practices include committee deliberations, appeal to the professional literature, and the employment of consultants. In a few areas, formal instruments exist to aid decision-makers in making input decisions. In the design of testing programs, for example, one may obtain substantial help by referring to the Buros' *Mental Measurements Yearbooks*.[14] Or in educational research, researchers who want to select an experimental design can receive material assistance in identifying and assessing alternative experimental designs by referring to the Campbell-Stanley chapter on experimental design in Gage's *Handbook on Research in Teaching*.[15] In this chapter, the decision situation posed the researcher in need of an experimental design is neatly laid out in the form of alternatives which are relevant to experimental research. Each of these designs is rated on the basis of its potential to meet criteria of internal and external validity as identified for each of the listed designs.

Decisions based upon input evaluation usually result in the specification of procedures, materials, facilities, schedule, staff requirements, and budgets in proposals to funding agencies. From the information provided in the proposals, the funding agencies in turn do an input evaluation to determine whether or not to fund the proposed projects. Funding agencies commonly employ expert consultants to serve as judges in their input evaluations.

Process Evaluation

Once a designed course of action has been approved and implementation of the design has begun, process evaluation is needed to provide periodic feedback to project managers and others responsible for continuous control and refinement of plans and procedures. The objective of process evaluation is to detect or predict, during the implementation stages, defects in the procedural design or its implementation. The overall strategy is to identify and monitor, on a continuous basis, the potential sources of failure in a project. These include, but are not limited to, interpersonal relationships among

staff and students; communication channels; logistics; understandings of and agreement with the intent of the program by persons involved in and affected by it; and adequacy of the resources, physical facilities, staff, and time schedule.

Process evaluation does not require control over assignment of subjects to treatments, nor that the treatments be held constant. Its purpose is to assist project personnel to make their decisions a bit more rational in their continual efforts to improve the quality of the program. Thus, under process evaluation, the evaluator accepts the program as it is and as it evolves and monitors the total situation by focusing the most sensitive and non-intervening data collection devices and techniques that he can obtain on the most crucial aspects of the project. Such evaluation is multivariate, and not all of the important variables can be specified before a project is initiated. The process evaluator focuses his attention on theoretically important variates, but he also remains alert to any unanticipated but significant events. Under process evaluation, information is collected, organized systematically, analyzed periodically, and reported as often as project personnel require such information, daily if necessary.

Thus, project decision-makers are not only provided with information needed for anticipating and overcoming procedural difficulties, but also with a record of process information to be used later for interpreting project outcomes.

Product Evaluation

The objective of product evaluation is to measure and interpret attainments, not only at the end of a project cycle but as often as necessary during the project term.

The method is to define operationally and to measure criteria associated with the objectives of the activity, to compare these measurements with predetermined absolute or relative standards, and to make rational interpretations of the outcomes using the recorded context, input, and process information. Criteria for product evaluation may be either **instrumental** or **consequential,** a distinction pointed out earlier by Scriven.[16] **Instrumental** criteria are related to program outcomes which contribute to the achievement of behavioral objectives. So for example, Clark and Guba have developed a taxonomy of instrumental objectives and associated criteria which are related to educational change.[17] **Consequential** criteria are primarily those pertaining to behavioral objectives. By way of example, Bloom's *Taxonomy of Educational Objectives*[18] is useful in the identification of consequential objectives.

In the change process, product evaluation provides information for deciding to continue, terminate, modify, or refocus a change activity, and for linking the activity to other phases of the change process. For example, a product evaluation of a program to develop after school study for students from disadvantaged homes might show that the development objectives have been satisfactorily achieved and that the developed innovation is ready to be diffused to other schools which need such an innovation.

THE STRUCTURE OF EVALUATION DESIGN

Once an evaluator has selected the type of evaluation appropriate to the kind of decision he intends to service, he must then develop a design to implement the evaluation. This is a difficult task, as the authors have already asserted, because few generalized designs exist which are adequate to meet evaluation needs. These designs must therefore be generated de novo. It should be noted, however, that it is possible to develop these designs for all kinds of evaluation, i.e., whether context, input, process, or product evaluations, by going through a series of identical steps. The checklist shown as Figure 4 is offered for this purpose. The structure proposed here has six major parts: focusing the evaluation, information collection, information organization, information analysis, information reporting, and administration. Each of these parts will be considered separately.

Focusing the Evaluation

The first element in designing an evaluation is that of focusing. The purpose of this step is to spell out the ends for the evaluation and to define policies within which the evaluation must be conducted. Specifically this element includes three steps. The first step is to identify and define the decision situations to be served. Given our present low state of knowledge about decision-making in education, this is a very difficult task. However, it is a crucial step

Focusing the Evaluation*

1. Define the decision situation(s) to be served, and describe each one in terms of its locus, criteria, decision rules, timing, and decision alternatives.
2. Define the system to be evaluated.
3. Define the evaluation specifications.

Collection of Information

1. Specify each item of information that is to be collected.
2. Specify the populations, sources, and sampling procedures for information collection.
3. Specify the instruments and methods for information collection.
4. Specify the arrangements and schedule for information collection.

Organization of Information

1. Specify a format for organizing the information.
2. Specify a means for coding, organizing, storing, and retrieving the information.

Analysis of Information

1. Specify the procedures for analyzing the information.
2. Specify a means for performing the analysis of information.

Reporting of Information

1. Specify the audiences for the evaluation reports.
2. Specify formats for the evaluation reports and reporting sessions.
3. Specify a means for providing the information to the audiences.
4. Specify a schedule for reporting the information to the specified audiences.

Administration of the Evaluation

1. Summarize the evaluation schedule.
2. Define staff and resource requirements and plans for meeting these requirements.
3. Specify means for meeting policy requirements for conduct of the evaluation.
4. Appraise the potential of the evaluation design for providing information which is valid, reliable, credible, timely, and pervasive.
5. Specify and schedule means for periodic updating of the evaluation design.
6. Provide a budget for the total evaluation program.

*The logical structure of evaluation design is the same for all types of evaluation, whether context, input, process, or product.

Figure 4: Developing Evaluation Designs

and should be carried out with as much rigor as practicable. Decision situations should be identified in terms of questions to be answered. Then one should identify those responsible for making the decisions, e.g., teachers, principals, board of education members, state legislators, and the like. Next, the criteria and decision rules to be employed should be identified. Then the timing of the decision situation to be served should be estimated so that the evaluation can be geared to provide relevant data prior to the time when the decision must be made. Finally, an attempt should be made to explicate each important decision situation in terms of the alternatives which might reasonably be considered.

Once the decision situations to be served have been explicated, the next step in the focusing activity is to define the setting within which the evaluation is to be conducted. Specifically, one should define the system in terms of its boundaries, its elements, and the characteristics of the elements. To return to an earlier example, the boundaries in a particular situation may be the inner city schools of a certain metropolitan area; the elements may be defined as, say, the pupils, the teachers, the parents and other patrons of the school, the program, the facilities, and similar elements; while the characteristics of a particular element, say the pupils, might be defined to include age, grade placement, intelligence score, sibling order, native language, and the like.

The third step in focusing the evaluation is to define policies within which the evaluation must operate. For example, one should determine whether a "self evaluation" or "outside evaluation" is needed. Also, it is necessary to determine who will receive evaluation reports and who will have access to them. Finally, it is necessary to define the limits of access to data for the evaluation team.

Collection of Information

The second major part of the structure of evaluation design is that of planning the collection of information. This section must obviously be keyed very closely to the criteria which were identified in the focusing step. So for example, if cost is a criterion factor, one must be sure to collect cost information.

Using those criteria one should first identify the sources of the information to be collected. These information sources should be defined in two respects: the origins for the information, e.g., students, teachers, principals, or parents, and the present state of the information, i.e., in recorded or non-recorded form.

Next, one should specify instruments and methods for collecting the needed information. Examples include achievement tests, interview schedules, and searches through the professional literature. Michael and Metfessel[19] have recently provided a comprehensive list of instruments with potential relevance for data collection in evaluations which will be helpful in this connection.

For each instrument that is to be administered, one should next specify the sampling procedure to be employed. Where possible, one should avoid administering too many instruments to the same person. Thus, sampling without replacement across instruments can be a useful technique. Also, where total test scores are not needed for each student, one might profitably use multiple matrix sampling where no student attempts more than a sample of the items in a test.

Finally, one should develop a master schedule for the collection of information. This schedule should detail the interrelations between samples, instruments, and dates for the collection of information.

Organization of Information

A frequent disclaimer in evaluation reports is that resources were inadequate to allow for processing all of the pertinent data. To avoid this problem, definite plans should be made regarding the third part of evaluation design: organization of information. Organizing the information that is to be collected includes providing a format for classifying information and designating means for coding, organizing, storing, and retrieving the information.

Analysis of Information

The fourth major part of evaluation design is analysis of information. The purpose of this part is to provide for the descriptive or statistical analyses of the information which is to be reported to decision-makers. This part also includes interpretations and recommendations. As with the organization of information, it is important that the evaluation design specify

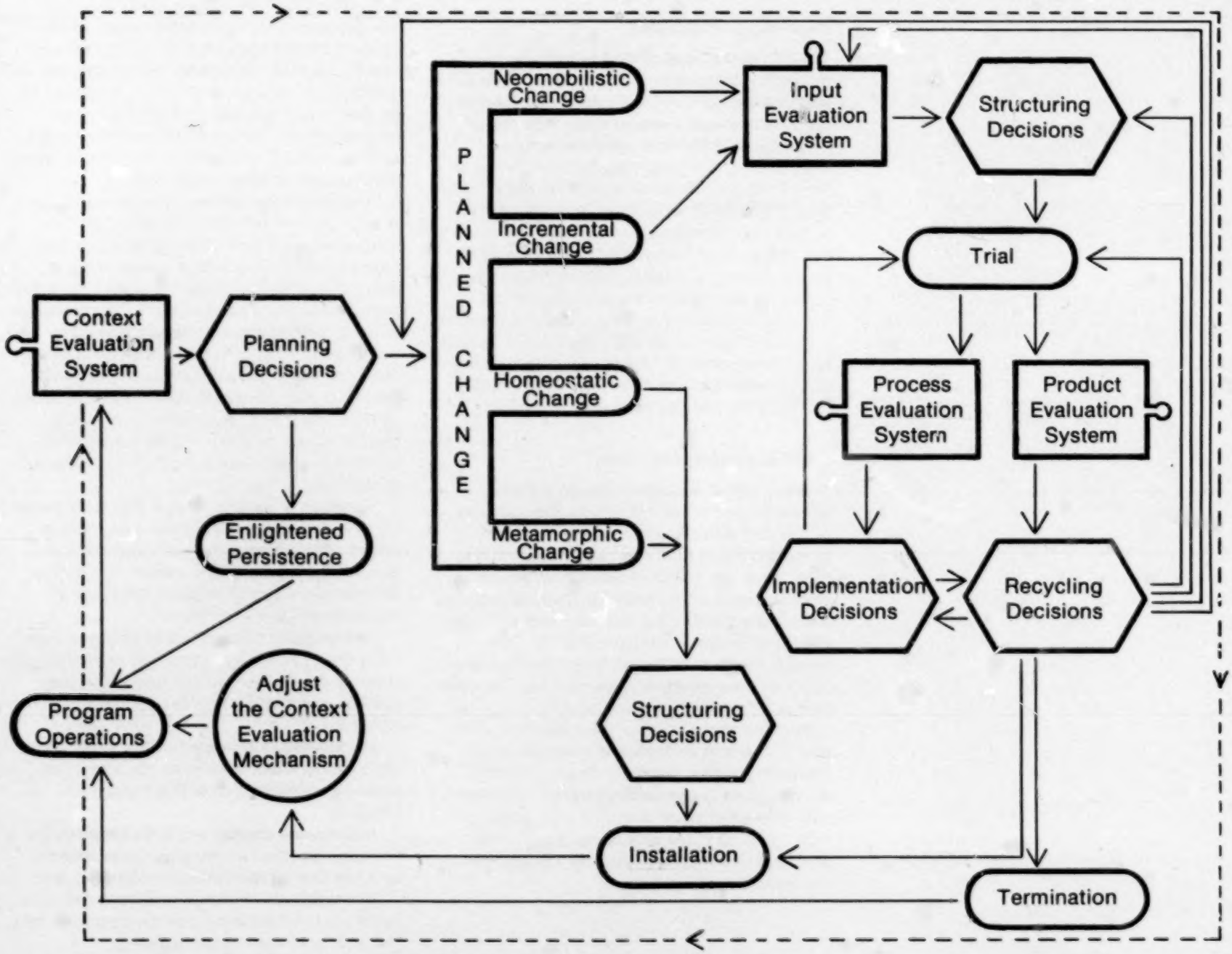

Figure 5: The CIPP Evaluation Model

means for performing the analyses. The role should be assigned specifically to a qualified member of the evaluation team or to an agency which specializes in doing data analyses. Also, it is important that those who will be responsible for the analysis of information participate in designing the analysis procedures.

Reporting of Information

The fifth part of evaluation design is the reporting of information. The purpose of this part of a design is to insure that decision-makers will have timely access to the information they need and that they will receive it in a manner and form which facilitates their use of the information. In accordance with the policy for the evaluation, audiences for evaluation reports should be identified and defined. Then means should be defined for providing information to each audience. Subsequently, the format for evaluation reports and reporting sessions should be specified. Finally, a master schedule of evaluation reporting should be provided. This schedule should define the interrelations between audiences, reports, and dates for reporting information.

Administration of Evaluation

The last part of evaluation design is that of administration of the evaluation. The purpose of this part is to provide an overall plan for executing the evaluation design. The first step is to define the overall evaluation schedule. For this purpose it often would be useful to employ a scheduling technique such as Program Evaluation and Review Technique (PERT). The second step is to define staff requirements and plans for meeting these requirements. The third step is to specify means for meeting policy requirements for conduct of the evaluation. The fourth step is to evaluate the potential of the evaluation design; criteria for such an evaluation is given in a following section. The fifth step is to specify and schedule means for periodic updating of the evaluating design. The sixth and final step is to provide a budget for the evaluation.

A TOTAL EVALUATION SYSTEM

Reliance upon ad hoc evaluation studies can prove to be an ineffective and inefficient means of providing information for decision-making within a system. Rather, educational systems should have well-functioning evaluation programs which provide a dynamic baseline of information about the system. Such an evaluation program should meet the regular, evaluative information requirements of the system, and it should be responsive to emergent needs for idiosyncratic data. Figure 5 is presented as an overview of the total evaluative program being proposed herein which provides for systematic context evaluation and ad hoc input, process, and product evaluations.

The outer loop represents a continuous, systematic context evaluation mechanism. This mechanism provides information to the planning body of a system for its use in making decisions either to change the system or to continue with present procedures in the knowledge that it is serving important objectives effectively and efficiently. If the context evaluation indicates that there are no discrepancies between the intentions and actualities or between possibilities and probabilities, the planning body likely would make choices which would result in a "steady-as-you-go" or "enlightened persistance" state.

However, if the context evaluation indicated that the program is deficient in some way, a rational decision-making body likely would decide to bring about changes to remove the deficiencies. Such changes as indicated in Figure 5 can be of four types.

Metamorphic change would be based upon decisions to effect large changes in the program supported by a high level of relevant understanding concerning how to bring about such changes.

Homeostatic change would be based upon decisions to effect small changes supported again by a high level of relevant understanding.

Incremental change would be based upon decisions to effect a small change supported by a low level of relevant understanding, and

Neomobilistic change would be based upon decisions to effect large change supported by a low level of relevant understanding.

Depending upon the type of change which results from planning decisions, vastly different

evaluation measures might be needed. In response to homeostatic or metamorphic change where adequate information to support decision-making is already available from the research literature and/or the context evaluation mechanism, it would be unwise to mount an expensive evaluation study to provide information which is redundant to that which already is available to the decision-maker. Therefore, our model shows: (1) that decision-makers would make structuring decisions regarding the means necessary to bring about homeostatic or metamorphic change **without** any intervening evaluation support mechanism, other than content evaluation, and (2) that these structuring decisions would lead **directly** to installation of change in the program and subsequent adjustment of the context evaluation mechanism so that the new feature in the system would routinely be monitored by the systematic context evaluation.

If, on the other hand, neomobilistic or incremental changes are called for, there is a definite need for ad hoc evaluation mechanisms to support such change, for both the context evaluation mechanism and the research literature provide inadequate supplies of information to support these types of changes.

First, an input evaluation study must be done to identify and evaluate strategies and procedures which could be used to effect desired changes. Such input evaluation information should assist decision-makers to make judgments in designing desired change procedures. In turn, the structuring decisions usually lead to some kind of a trial or pilot phase, for as yet, the desired change is an innovation, and has not been adequately tested. It is, therefore, not ready for installation in the total system.

Process and product evaluation are next included to aid in decisions pertaining to the trial phase. Process evaluation would provide information for implementation decisions needed for efficient operation of the trial, including the recycling of structuring decisions as necessary. Product evaluation would go on simultaneously throughout the process of the trial in conjunction with process evaluation and would support recycling decisions which could lead to a reformulation of the change to be brought about, a modification either in strategy or procedure, termination of the change effort, or in the installation of the innovation in the total program. In the case of installation, again, the context evaluation mechanism would be adjusted to allow systematic monitoring of the new element in the total system.

CRITERIA FOR JUDGING EVALUATIONS

How can the evaluator evaluate his own activity? The information which the evaluation produces is the key. What criteria are appropriate to it?

This question can be answered in two parts. If evaluation produces information, then that information must meet criteria that are ordinarily required of any good information, i.e., scientific criteria. But because it is evaluative information, it must also meet certain special criteria of practical utility.

The scientific criteria are these:

Internal validity The information provided by the evaluation must display a reasonable correspondence to the phenomena which it purports to describe or interpret. It must have fidelity, or, in the layman's sense, it must be true.

External validity The information must be generalizable to similar situations beyond the one in which it was collected. Particularistic data have little utility. If, for example, data relating to the effectiveness of an innovation could not be interpreted as also being valid in classrooms other than the ones in which they were collected, little would be gained in deciding whether to adopt or not.

Reliability Here the concern is with the replicability of the data. If a repetition of the evaluation did not produce essentially similar findings, we should be concerned that the findings were simply random and therefore meaningless.

Objectivity Here concern is with the publicness of the data. If data are private in the sense that only particular persons would so interpret them, i.e., that not all competent judges would agree on them, their true meaning is subject to question.

In addition to these four general criteria that could be invoked in relation to any information, certain special criteria of practical utility must be met by evaluative information. These are:

Relevance The information must relate to the decisions to be made.

Significance The information must be weighted for its meaning in relation to the decision. Not all relevant information is equally weighty. The culling and highlighting required is a professional task that justifies the inclusion of a reportorial expert on the evaluation team.

Scope The information must relate to all aspects involved in the decision. If there are six alternatives to be considered, information that applies to only four lacks scope.

Credibility The information must be trusted by the decision-maker.

Timeliness The information must come in time to be useful to the decision-maker. The evaluator must guard against the scientific value that argues against publishing findings until every last element is in. Late information is worthless information. It is better in the evaluative situation to have reasonably good information on time than perfect information too late.

Pervasiveness The information must get to all of the audiences (i.e., to all of the decision-makers) who need it.

Efficiency It is possible for an evaluation to mushroom out of all proportions to its value. The imprudent evaluator may produce a mountain of information whose collection imposes an intolerable financial drain. Proper application of the criteria of relevance, significance, and scope should remedy the grossest inefficiencies. But even when the information proposed to be collected meets all of these criteria, there are probably still alternative ways for collecting it that differ in terms of the time, costs, personnel, etc., that are required. The criterion of efficiency will guide the evaluator to the appropriate alternative.

An evaluator who can say, after careful examination, that his evaluation design will produce information that conforms to all of these criteria can be assured that he is doing his job well.

Footnotes

[10]*Webster's New World Dictionary*, College Edition, New York: World Publishing Company, 1966, p. 749.

[11]Lee J. Cronbach. "Course Improvement Through Evaluation," *Teachers College Record*, 64 (May 1963), p. 672.

[12]David Braybrooke and Charles E. Lindblom. *A Strategy of Decision*, New York: The Free Press, 1963. This most insightful work is based on the authors' wide experience in the arena of public policy decision-making. A significant portion of the book deals with a discussion of the inadequacies of ordinary formulations of decision-making processes, which treat decision-making as rational. Braybrooke and Lindblom instead espouse a strategy of decision-making which they term "disjointed incrementalism," and which is based on the lower left quadrant of Figure 2 (q.v.), called "Incrementalism" by the present authors. The educational situation is sufficiently different from that normally encountered in public policy arenas to make viable certain decision strategies in the upper left and lower right quadrants of Figure 2 (homeostasis and neomobilism), which quadrants Braybrooke and Lindblom believe have little utility in guiding policy decision strategies in most cases. It is chiefly in this regard that the formulations in this paper differ from Braybrooke and Lindblom. The present authors acknowledge a great indebtedness to them for the concepts of high vs. low change and high vs. low understanding, which form the basis for the strategies implied in Figure 2 and which are further explicated in this text.

[13]*Ibid*, p. 66.

[14]Oscar K. Buros. *Mental Measurement Yearbook*, Highland Park, New Jersey: Gryphon Press, 1938, 1940, 1949, 1953, 1959, and 1965.

[15]N.L. Gage, Editor. *Handbook on Research in Teaching*, The American Educational Research Association, Chicago: Rand McNally and Company, 1963.

[16]Michael Scriven. *The Methodology of Evaluation*, Bloomington, Indiana: Indiana University, Social Science Education Consortium, Publication #110, 1965.

[17]David L. Clark and Egon G. Guba. "An Examination of Potential Change Roles in Education." Paper read at the Symposium on Innovation in Planning School Curricula, Airlie House, Virginia, October 1965.

[18]Benjamin S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*, New York: Longmans, Green and Company, Inc., 1956.

[19]Newton S. Metfessel and William B. Michael. "A Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs." *Educational and Psychological Measurement*, 1967, 27, pp. 931-936.